NTP
National Toxicology Program

Levels of Evidence Criteria for NTP Developmental Toxicology Studies

Edward Carney, Ph.D.
Chair, Reproductive and Developmental Toxicity Criteria Working Group

NTP Board of Scientific Counselors Meeting

November 20-21, 2008

Introductory Comments

- Same key points as for reproductive toxicology studies:
  - Applies to individual studies of chemical agents
  - Notes the strength of the evidence
  - Negative results only imply that the chemical is not a developmental toxicant under the specific conditions of the study
  - Positive results are assumed to be relevant to humans, unless data are available which demonstrate otherwise.
  - Developmental events are intertwined in the reproductive process. Effects on developmental toxicity may be detected in reproductive studies
  - Communication: hazard only; study results do not imply risk!
  - Five categories
  - Report LOEL (clear, some evidence) or NOEL (no evidence)

Levels of Evidence for Developmental Toxicity - 1

- **Clear Evidence of Developmental Toxicity**
  - Demonstrated by a dose-related[1] effect on one or more of its four elements (embryo-fetal death, structural malformations, growth retardation or functional deficits) that is not secondary to excessive maternal toxicity. A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for developmental toxicity" should accompany the evidence statement.

  - [1]The term "dose-related" describes any dose relationship, recognizing that the treatment-related responses for some endpoints may be non-monotonic due to saturation of exposure or effect, overlapping dose-response behaviors, change in manifestation of the effect at different dose levels, or other phenomena.

Levels of Evidence for Developmental Toxicity - 2

Some Evidence of Developmental Toxicity

- **Some evidence** of developmental toxicity, relative to clear evidence, is characterized by greater uncertainties or weaker relationships with regard to dose, severity, magnitude, incidence, persistence, and/or decreased concordance among affected end points.

- A statement to the effect of "This study has a lowest observed adverse effect level of XXXX mg/kg/d for developmental toxicity" should accompany the evidence statement, except in those instances in which the "some" classification has been based on uncertainties about the dose relationship that precludes confident determination of the LOAEL.

Levels of Evidence for Developmental Toxicity - 3

- **Equivocal Evidence of Developmental Toxicity**
  - Demonstrated by marginal or discordant effects on developmental parameters that may or may not be related to the test article.
- **No Evidence of Developmental Toxicity**
  - Demonstrated by data from a well conducted, adequate study that are interpreted as showing no biologically relevant evidence of chemically-related effects on development. A statement to the effect of "This study had no observable adverse developmental toxicity at the highest dose tested (XXXX mg/kg/d)".
- **Inadequate Study of Developmental Toxicity**
  - Demonstrated by a study that, because of major design or performance flaws, cannot be used to determine the presence of developmental toxicity.

Key points to consider with the Levels of Evidence criteria

- When a conclusion statement for a particular experiment is selected, consideration must be given to key factors that would extend the boundary of an individual category of evidence. Such consideration should allow for incorporation of scientific experience and current understanding of developmental toxicity studies in laboratory animals, particularly with respect to interrelationships between end points,
  - impact of the change on developmental function,
  - relative sensitivity of end points, normal background incidence, and specificity of the effect.
- For those evaluations that may be on the borderline between two adjacent levels, some factors to consider in selecting the level of evidence of developmental toxicity are given below:

- Increases in severity and/or prevalence (more individuals and/or more litters) as a function of dose generally strengthen the level of evidence, keeping in mind that the specific manifestation may change with increasing dose. For example, malformations may be observed at a lower dose level, but higher doses may produce embryo/fetal death.

Other Key Points -2

- Because of the relationship between maternal physiology and development, evidence for developmental toxicity may be greater for a selective effect on the embryo-fetus or pup, although there may be exceptions.

- Effects seen in many litters may provide stronger evidence than effects confined to one or a few litters even if the incidence within those litters is high.

- Concordant effects (syndromic) may strengthen the evidence of developmental toxicity. Single endpoint changes by themselves may be weaker indicators of effect than concordant effects on multiple endpoints related by a common mechanism.

- In order to be assigned a level of "clear evidence" the endpoint(s) evaluated should normally show a statistical increase in the deficit, or syndrome, on a litter basis.

- In general, the more animals affected, the stronger the evidence; however, effects in a small number of animals across multiple, related endpoints should not be discounted, even in the absence of statistical significance for the individual endpoint(s). In addition, rare malformations with low incidence should be interpreted in the context of historical controls and may be biologically important.

Other Key Points -3

- Consistency of effects across generations in a multi-generational study strengthens the level of evidence. However, if effects are observed in the F_1 generation but not in the F_2 generation (or the effects occur at a lesser frequency in the F_2 generation), this may be due to survivor selection (i.e., if the effect is incompatible with successful reproduction, then the affected individuals will not produce offspring).

- Transient changes (e.g., pup weight decrements, reduced ossification in fetuses) by themselves may be weaker indicators of an effect than persistent changes.

- Insights from supportive studies (e.g., toxicokinetics, ADME, computational models, structure-activity relationships) and developmental findings from other *in vivo* animal studies (NTP or otherwise) should be drawn upon when interpreting the biological plausibility of an effect.

- Uncertainty about the presence of developmental toxicity in one study may be lessened by effects (even if not identical) that are observed in a second species.

Other Key Points - 4

- The studies should be well designed and be of adequate experimental design and statistical power.
- New technical approaches and highly sensitive techniques need to be appropriately characterized to build confidence in their utility, and their usefulness as indicators of effect is increased if they can be associated with changes in traditional endpoints.

QUESTIONS?